Comments on L. A. Paul, *Transformative Experience*, Oxford University Press, 2014.

Dilip Ninan
dilip.ninan@tufts.edu

CSMN Workshop on the First-Person Perspective
Norwegian Institute, Athens, Greece
May 2015

1. Introduction

Two ways in which an experience can be transformative:

- An **epistemically transformative** experience is an experience of a type that you've never had before; by undergoing the experience you learn 'what it is like' to have an experience of that type.

  Examples:

  - Eating something (e.g. a durian) you've never tasted before. Prior to eating a durian, you don't know what it is like to eat one. After trying it, you do.
  - Seeing a color you've never seen before.
  - Having children.
  - Becoming a vampire.
  - Gaining a new sensory ability.

- A **personally transformative** experience is an experience that alters you in a fundamental way; in particular, it can change your preferences and values in a radical way.

  Examples:

  - Experiencing a horrific physical attack.
  - Undergoing a religious conversion.
  - Having children.
  - Becoming a vampire.
  - Gaining a new sensory ability.

- An experience can be epistemically transformative without being personally transformative. For example: trying an exotic fruit may be epistemically transformative, but will probably not be personally transformative for most of us.

- Paul is particularly interested in those experiences that are transformative in both senses. These are what she calls *transformative experiences* (simpliciter).

Epistemically transformative choice:

- An **epistemically transformative choice** is a choice of whether to undergo an epistemically transformative experience.

  Examples:

  - Eat a durian or eat a pineapple.
  - Become a vampire or remain a human.
  - Have children or remain childless.

Ordinary decisions and normative decision theory:

> "Ordinary decision-making at the personal level, to be rational, must follow the rules of a realistic normative decision theory." 19
>
> "...the best choice for the normatively rational decision-maker is to choose to perform the act with the highest expected value, given her assignments of values to outcomes and probabilities to states..." 22

Two problems:

- **The problem of epistemically transformative choice**: When faced with an epistemically transformative choice, the agent can't rationally assign utilities to some of the relevant outcomes.
- **The problem of personally transformative choice**: When faced with a personally transformative choice, the agent's post-choice utility function may differ from her pre-choice utility function. So how should the agent decide?
- I'm going to discuss the first problem.

My comments:

- **First point**: It seems that, intuitively speaking, we can often make rational decisions even if we don't *know* our utilities for certain outcomes. We do this by *estimating* our utilities for certain outcomes, i.e. by going via are judgments about how likely it is that we will like something.
- **Second point**: Standard decision theory seems capable of taking the previous point onboard, though it requires a philosophically interesting modification.
- I don't know to what extent Paul disagrees with these points, but they will hopefully serve as useful starting points for discussion.

2. The problem of epistemically transformative choice

One version of normative decision theory ('counterfactual decision theory'):

- A formal decision problem consists of a set of possibilities W, a set of states $S = \{S_1, ..., S_n\}$, and a set of acts $A = \{A_1, .., A_k\}$, where A and S determine partitions of W.
- The **expected utility** of an act A_i is calculated as follows:

$$EU(A_i) = P(A_i \mathbin{\square\!\!\rightarrow} S_1)u(A_i \wedge S_1) + ... + P(A_i \mathbin{\square\!\!\rightarrow} S_n)u(A_i \wedge S_n) \quad (1)$$

- Here $\mathbin{\square\!\!\rightarrow}$ is the counterfactual conditional, P represents the credence function of the agent facing the decision, and u is the agent's utility function.
- If the agent chooses to perform act A_i, her choice is irrational if there is another act A_j whose expected utility is greater than that of A_i; otherwise, it is rational.

Paul's examples:

- Normally, in the sorts of decision problems discussed in decision theory, the agent is ignorant of some state of the world, e.g. whether it is going to rain, whether the Predictor put a million dollars in the opaque box, etc.
- But in many of Paul's examples, there doesn't appear to be any relevant ignorance of the state of the world at issue (a point to which I'll return).

- For example, suppose I'm deciding whether or not to eat a durian. I know (let us suppose) that it is ripe, that it won't cause me to be sick, etc.
- In that case, my decision problem will look like this:

  W, $S = \{W\}$, $A = \{$*I eat a durian*, *I don't eat a durian*$\}$

- In such cases, the expected utility of performing an action (e.g. eating a durian) is simply the utility of performing that action.[1]

The problem of epistemically transformative choice:

- One problem raised by epistemically transformative choices is that I don't know how to assess the utility of certain outcomes (an outcome is here understood as the conjunction of an act and a state).
- Since I've never eaten a durian before, I don't know what it's like to eat a durian, and don't know whether I'd enjoy the experience or not.
- Thus, I don't know what my utility for the outcome *I eat a durian* is.
- So I don't know what the expected utility of eating a durian is, and so I can't compare the expected utility of that action with my alternative, not eating the durian.
- So decision theory doesn't yield an answer to the question: what should I do?

> "To apply a normative decision-theoretic model... to a decision about whether to perform an act, you need to know the values of the relevant outcomes... But in the case of a decision involving a transformative experience, you cannot know what it is like to have that kind of experience until you have it. In this situation, you cannot determine the subjective value of the any outcome that involves what it is like for you to have... the experience. And if you cannot determine the subjective values of the outcomes, you cannot compare the values." 32

Justified assignments:

- Of course, I could just go ahead and assign these outcomes utilities. Or I might just find myself with certain brute preferences.
- Paul's problem with this, I take it, is that such an assignment would be unjustified:

> "If we are to meet the normative standard when we make our choices, we must be rationally justified in our assignments of values and credences to the outcomes and states of our decision problem. That is, we must assign our values and credences based on sufficient evidence. If we assign values and credences based on insufficient evidence and calculate the expected value of our acts using such assignments, our decision does not meet the normative standard for rationality. The lunatic in the asylum who see ghosts in every corner and assigns the values and credences on that basis does not meet the normative standard..." 22-23

- As I understand her, Paul is assuming that there is a fact of the matter as to (e.g.) what my utility for eating durian is, a fact which I might be ignorant of.

3. Beliefs about utilities

First point: At least in some cases, one can (intuitively speaking) make a rational decision even if one doesn't know what utilities to assign to some of the relevant outcomes.

[1] Since $P(A \mathrel{\square\!\!\rightarrow} W) = 1$ and $u(A \wedge W) = u(A)$.

- Suppose I've never tried a durian nor a rambutan.
- Still, it might be that I'm more certain than not that I'd like the taste of durian, and more certain than not that I would *not* like the taste of rambutan.
- But suppose my evidence as to whether or not I'd like these fruits is limited, so that I don't *know* whether or not I'd like them, nor do I know how much (or how little) I would like them.
- So I can't really assign utilities to the relevant outcomes, since I don't kow what the relevant utilities are.
- Still if these are the only two fruits at the breakfast buffet, it seems entirely reasonable for me to choose the durian rather than the rambutan (suppose that I must choose one).
- What appears to make this rational is my *estimation* of what my (unknown) utilities are. If my belief that I will probably like durian and my belief that I will probably not like rambutan are justified, then that seems to justify my choice.

Second point: Decision theory can be modified to accommodate this fact.

Refining the decision problem:

- Consider again the case in which I face the choice of eating a durian or refraining from eating it.
- Originally, we said that the decision problem I faced could be represented as follows:

  W, $S = \{W\}$, $A = \{$*I eat a durian*, *I don't eat a durian*$\}$
- Here, the outcomes are *I eat a durian* and *I don't eat a durian*. We agreed that we couldn't assign utilities to the first outcome, since I don't know what that outcome is like, and I don't know whether or not I'd like it.
- Since (we can suppose that) most people either love durian or hate it, let's say there are two relevant utility functions that, for all I know, might be mine, u_1 and u_2, with u_1 assigning a high utility to eating durian (say 10), u_2 assigning a low utility to eating it (say -10).
- Our original model of the decision problem doesn't represent the fact that I'm uncertain about a crucial piece of information: whether my utility function is u_1 or u_2.
- We can bring this into the model by refining the states in our decision problem as follows (letting U_n be the proposition that my utility function is u_n):

  W, $S = \{U_1, U_2\}$, $A = \{$*I eat a durian*, *I don't eat a durian*$\}$

Expected utility again:

- Now, to define the expected utility of an action, we may need a sort of meta-utility function, one that assigns utilities to possible states of the world that themselves incorporate information concerning what my utility function is.
- Call this meta-utility function v, and assume that $v(A \wedge S \wedge U_n) = u_n(A \wedge S)$. That is, the value I assign to being in a situation in which A and S obtains and in which my utility function is u_n is simply whatever u_n assigns to $A \wedge S$.
- The formula for expected utility is as before, but now v replaces u.
- There is some work to be done in thinking through what (if anything) this is saying about the psychology of rational agents. It might be interpreted as building in an assumption about how my higher-order desires related to my first-order desires.

Applied to the durian case:

- Given the set-up of the durian case, the expected utility of D, *I eat a durian*, will be determined as follows:[2]

$$\begin{aligned} EU(D) &= P(U_1)v(D \wedge U_1) + P(U_2)v(D \wedge U_2) \\ &= P(U_1)u_1(D) + P(U_2)u_2(D) \end{aligned} \tag{2}$$

- But here the values of $u_1(D)$ and $u_2(D)$ are, by hypothesis, known: $u_1(D) = 10$ and $u_2(D) = -10$.
- Suppose the expected utility of refraining from eating a durian is 0.
- Then I should eat the durian if it is more probable than not that my utility function is u_1, i.e. it if is more probable than not that I will like eating a durian.
- Thus, even if I don't know what utility to assign to *I eat a durian*, as long as I can rationally assign credences to propositions of the form *my utility for eating a durian is n*, decision theory can be used to evaluate my choice in this case.

Comments:

- Even if my suggestion on behalf of decision theory works, I still think we are likely to learn something about decision-making from considering Paul's problem of epistemically transformative choice (or from the more general problem of unknown utilities).
- The issue may have consequences for how we interpret formal theories of decision.
- I haven't said anything about the problem of *personally transformative choice*, the problem that arises from the fact that a personally transformative experience might change one's preferences.

[2] I'm assuming here that the issue of whether my utility function is u_1 or u_2 is counterfactually independent of my eating the durian, so that (e.g.) $P(D \mathbin{\square\!\!\rightarrow} U_1) = P(U_1)$.